Description

FOWLER-NORDHEIM BLOCK ALTERABLE EEPROM MEMORY CELL

TECHNICAL FIELD

The present invention relates in general to semiconductor devices. More specifically, the present invention relates to block alterable memory devices.

BACKGROUND ART

The need for a high-density block alterable memory ~~device~~ devices is ever increasing. Cellular phones, memory sticks, digital cameras, laptop computers, and ~~palm pilots~~ personal data assistants are a few examples of small devices that demand higher density memories. These devices require alterable memories because their contents change every time they are in use. For example, the size of a memory stick is as small as a pen but it can store 256 MB memory. The memory stick has a Universal Standard ~~Board~~ Bus (USB) port that can plug into another USB memory port of a computer to transfer the data from the hard drive of that computer. Therefore, the memory stick and other similar devices such as camera memories need a high-density alterable memory device to erase old data and store new data. The Electrically erasable programmable read only memory (EEPROM) common in the industry cannot be used in these applications because EEPROM ~~cannot be~~ is not alterable under normal operation conditions.

A typical block alterable memory device employs flash memory to program, read, or erase memory cells. With reference to Fig. 1, a flash memory 100 is a memory array which is arranged in rows 102 and columns 106. Each row 102 has N+1 memory cells connecting to source lines S_0 - S_N. The first memory cell in the row 102

belongs to column BL_0 and the N^{th} memory cell belongs column BL_N. Therefore, there are N+1 columns in the flash memory array 100. The gates of all the cells within a ~~row~~ column 106 are coupled together to form a wordline WL_i 102. There are M+1 wordlines or rows in the flash memory array 100, ranging from WL_0 to WL_M. The sources of the cells in each column are coupled together and coupled to the select lines 104, ranging from S_0 to S_N. The drains of the cells in each row are coupled together to form a bitline 106, ranging from BL_0 to BL_N. The flash array 100 enables users to electrically program and erase information stored in a memory cell 110.

Each memory cell 110 in the flash memory matrix 100 is a floating gate transistor. The structure of a floating gate transistor is similar to a traditional MOS device, except that an extra ~~poly-silicon~~ polysilicon strip is inserted between the gate and the channel. This strip is not connected to anything and called a floating gate. The threshold voltage of a floating gate transistor is programmable. The described flash memory 100 uses the ~~Fowler Nordheim~~ Fowler-Nordheim tunneling effect to program a cell 110. Programming is a process ~~when~~ wherein electrons are ~~place~~ placed in the floating gate. Programming occurs when applying a high voltage between the gate[[,]] and source, and ~~gate-drain~~ gate and drain terminals that a high electric field causes injection of carriers into the floating gate. Electrons acquire sufficient energy to become hot and traverse ~~through~~ the first oxide insulator, so they get trapped on the floating gate. Programming is done on a bit basis by applying a correct voltage at the bitline 106 of each cell 110.

The floating gate layer allows the cell 110 to be electrically erased through the gate. Erase and program operations of the memory array 100 can be done on

more than one cell at a time. However, the alterable flash memory device has reliability and durability problems because the voltages for erasing[[,]] and programming are very high.

With reference to Table 1 at the end of this specification, in order to achieve block alterable memory, the memory cell 110 in the flash memory array 100 as shown in Fig. 1 needs to apply +10 volts or -10 volts across the wordline WL_i 102, the source line S_i 104, and the bitline BL_i 106. Accordingly, the placement of such high voltages to a single memory cell transistor 110 presents reliability and durability problems. Over long periods of time, placing high voltages on the memory device 100 may alter a program stored in each cell 110.

One prior art solution to this problem (for example, U.S. Patent No. 5,066,992 to T.C. Wu) is shown in Fig. 2A. This solution places an extra select transistor 202A in series with a flash memory cell 210A. The gate of the additional select transistor 202A is coupled to the select line S_0 to S_N, the drain 204A is coupled to the bitline BL_0 to BL_N 214A, and the source 206A is coupled to the drain of the flash cell. Thus, when a select line S_i is ON, each selected transistor connected to the select line S_i is turned ON. As a result, the voltage of the drain of the flash cell 210A is proportional to the voltage of the bitline B_i. During a reading cycle, the bitline 214A is open, the select line S_i is grounded, and the wordline WL_i is at negative program voltage V_D. Thus, a program stored in an EEPROM device 200A remains unaltered. Thus, the memory array ~~200~~ 100 lasts longer and avoids the reliability and durability of one-transistor memory cells presented above. However, the two-transistor memory cells require larger areas for manufacturing because each memory cell has two transistors.

Referring to Fig. [[2]] 2B, ~~a plan view and~~ various cross-sectional views of a memory array 200B are shown. Memory array 200B is formed on a face of a semiconductor substrate 222B. Substrate 222B ~~exhibits~~ is doped with a p-type majority carrier. Bitline BL 214B, select line [[S_i]] SEL ~~202~~ 202B, wordline [[WL_i]] WL 208B, and the source are n-type and implanted within substrate 222B at the surface. The gate 208B comprises [[of]] a first poly layer 209B, a second poly layer 211B, and an inter poly layer 212B. Accordingly, column lines 214B and 206B serves as a source and drain[[s]] of transistors which are used in forming memory cells contained within memory array 200B. Each of the column lines 214B serves as a source of one memory cell or a drain of an adjacent cell. However, this solution dedicates large sections on the semiconductor substrate to the alterable block function. An undesirably low density flash memory results. Consequently, the industry has a need for a memory device structure which has block alterable capability without dedicating semiconductor substrate area to that function.

U.S. Patent No. 4,783,766 to Samachisa et al. describes [[an]] a memory cell of a block alterable EEPROM in which a single control gate is common to both the floating gate memory cell and the select transistor device. However, the device is formed using a different process flow from that of flash memory devices, thus requiring a separate ~~mask~~ masking sequence.

U.S. Patent No. 6,420,753 to Hoang describes a similar structure to that of the Samachisa patent. It is stated that these memory cells can be manufactured without requiring additional processing steps from those of comparable flash memories.

SUMMARY OF THE INVENTION

A ~~Fowler-Norheim~~ Fowler-Nordheim block alterable memory cell in accordance [[to]] with the present invention is carried out in one form by a memory cell constructed from two separate transistor cells that have common ~~select control~~ select-control gate. The two cells are constructed on a substrate or in a well that exhibits a first (e.g., [[p]] "p" or acceptor) conductivity type. A tunnel oxide layer resides on the substrate face. The ~~select control~~ select-control gate comprises a first poly layer, an interpoly layer, and a second poly layer. The second poly layer is extended to connect to the gate of the first cell to form a common ~~select control~~ select-control layer. The extended portion of the common ~~select control~~ select-control layer contacts a drain implant region. A buried n+ implant region is formed near the surface of the p-substrate. The floating gate region is positioned above the buried implant and extends over the channel of transistor 400B. A self-aligned source/drain implant is located at edges of the control poly. The area of the substrate between the floating gate region and the drain implant region that lies underneath the extended portion of the common ~~select control~~ select-control layer is known as the active region. Thus, the ~~Fowler-Norheim~~ Fowler-Nordheim block alterable memory device in accordance [[to]] with the present invention is constructed as a single transistor memory cell but it behaves as a two transistor[[s]] cell because of the extended ~~select control~~ select-control layer.

~~In another aspect of the~~ The present invention is also a method to fabricate a memory cell~~, in order~~ to achieve a ~~Fowler-Norheim~~ Fowler-Nordheim block alterable memory cell as described above~~, the memory cell is manufactured according to a method~~. The method first deposits a screen oxide of ~~thickness~~ about 150 angstroms

thickness over the ~~p-substrate~~ p-type substrate. Then a photoresist mask with an opening is added on top of the screen oxide layer. ~~Cell~~ A cell channel ~~implant~~ and buried n^+ ~~implant~~ implants are implanted at the location of the opening of the mask and near the surface of the p-substrate. Next, ~~the method etches out~~ the screen oxide is etched and ~~grows~~ initial gate oxides are grown. A tunnel window mask is then formed. A tunnel oxide is etched in the screen oxide layer where the windows of the tunnel window mask are located. The first polycrystalline silicon (poly) layer over the tunnel oxide and cell implants are deposited. An insulating layer is formed overlying the first poly layer. An extended final (second) poly layer is deposited over the insulating layer. Finally the ~~method~~ device is completed by source and drain ~~implant~~ implants.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 illustrates a schematic diagram of a prior art memory array having [[a]] single flash memory ~~cell~~ cells.

Fig. 2A illustrates a schematic diagram of a prior art dual transistor memory cell that has block alterable capability. The top cell is used to define the block to be altered, and the second cell or flash cell is used to store data information.

Fig. 2B illustrates a sectional view of the dual transistor memory cell of Fig. 2A.

Fig. 3 illustrates a schematic diagram of a ~~Fowler-Norheim~~ Fowler-Nordheim alterable block memory array in accordance with the present invention.

Fig. 4A illustrates a schematic diagram of a single cell from the ~~Fowler-Norheim~~ Fowler-Nordheim alterable block memory array in accordance with the present invention.

Fig. 4B illustrates a cross sectional view of a ~~Fowler-Norheim~~ Fowler-Nordheim block alterable memory cell as illustrated in Fig. 4A.

~~Fig. 5A~~ Figs. 5A-5G illustrate[[s]] exemplary cross sectional views of process steps in accordance with the present invention ~~a substrate with screen oxide layer on top in accordance to step one of the method of the invention~~.

Fig. 5A illustrates a cross sectional view of a substrate with a screen oxide layer.

Fig. 5B illustrates a cross sectional view of a barrier mask layer ~~and the~~ with a window ~~in the middle~~ for depositing a cell implant in the substrate ~~according to step 2 of the invention~~.

Fig. 5C illustrates a cross sectional view of a ~~mask out~~ mask-out implant ~~according to step 3 of the invention~~.

Fig. 5D illustrates a cross sectional view of a p-substrate with a buried n+ implant, a source implant, and a drain implant.

Fig. 5E illustrates a cross sectional view of a ~~Fowler-Norheim~~ Fowler-Nordheim cell with a window tunnel mask and etch oxide ~~according to step 4 of the present invention~~.

Fig. 5F illustrates a cross sectional view of a ~~Fowler Norheim~~ Fowler-Nordheim block alterable memory cell with a tunnel oxide layer and a first polycrystalline layer ~~according to step 5 of the present invention~~.

Fig. 5G illustrates cross sectional view of a ~~Fowler Norheim~~ Fowler-Nordheim block alterable cell with an oxide-nitride-oxide (ONO) deposition and a ~~the~~ control poly layer deposition ~~according to step 6 of the present invention~~.

Fig. 6 illustrates a flowchart of ~~the~~ a method for manufacturing the ~~Fowler-Norheim~~ Fowler-Nordheim block alterable cell corresponding to Figs. ~~4A-4G~~ 5A-5G.

~~PREFERRED EMBODIMENT OF THE DESCRIPTION~~
BEST MODE FOR CARRYING OUT THE INVENTION

Figs. 3, 4A, and 4B show various views of ~~a preferred~~ an exemplary embodiment of a ~~Fowler-Norheim~~ Fowler-Nordheim block alterable memory architecture fabricated according to ~~the~~ a method of the present invention. With respect to Fig. 3, ~~shows an array of cells of~~ a memory array 300[[,]] includes a plurality of memory cells 310. The memory cell 310 includes ~~in which~~ a select ~~control~~ transistor 302 and a ~~flash~~ memory transistor 304 ~~with~~ which share a common control gate ~~form a single memory cell 310~~. The plurality of memory cells 310 may be erased and programmed in blocks or programmed or read bit-by-bit by applying appropriate voltages to the bitlines ~~(B_0 to B_N)~~ (BL_0 to BL_N), source lines (S_0 to S_N) and wordlines (WL_0 to WL_M). Typically all memory cells 310 in the array 300 are normally constructed as a result of ~~the same~~ similar process steps, and therefore, all cells are similar in structure.

Referring to Fig. 4A, a schematic of a memory cell ~~400A~~ 310 in accordance [[to]] with the present invention ~~is shown. Each memory cell 400A~~ includes a memory transistor ~~404~~ 304 connected in series with a select transistor ~~402~~ 302 at ~~the~~ a drain/source junction. (The drain of the memory transistor ~~404~~ 304 is coupled to the source of the select transistor ~~402~~ 302.) The source of the ~~flash cell 404~~ memory transistor 304 is coupled to a select line S_i. The drain of the ~~control~~ select transistor 302 is coupled to a bitline BL_i. ~~Their~~ A common gate of the memory transistor 304 and select transistor 302 is coupled to a wordline WL_j. This common gate for the ~~two~~ memory 304 and select 302 transistors

~~402 and 404~~ can be ~~manufacturing~~ manufactured as a single cell having an extended and continuous poly layer, thus reducing ~~the~~ an area of the memory cell 310 ~~cell's area~~.

With reference to Fig. 4B, a cross sectional view 400 of ~~a single~~ the memory cell ~~400B~~ 310 is illustrated. The memory cell ~~400B~~ 400 is formed on a semiconductor substrate (or well) ~~401B~~ 401 of a first conductivity type, which in the ~~preferred~~ exemplary embodiment is p-type. A drain implant region ~~402B~~ 402 and a source implant region ~~406B~~ 406, ~~respectively,~~ are implanted within an uppermost surface of the substrate ~~401B~~ 401 ~~near the surface~~. A buried, heavily doped implant ~~404B~~ 404 for the floating gate region is also formed within an uppermost surface of the substrate ~~401B~~ 401. The implant regions ~~402B, 404B and 406B~~ 402, 404, and 406 are of a second conductivity type of a polarity opposite that of the conductivity type of the ~~exhibited by~~ substrate ~~401B~~ 401. In a specific exemplary ~~preferred~~ embodiment, the implants are ~~n type~~ n-type. The [[n^+]] buried implant ~~404B~~ 404 is of n+ conductivity and serves as a tunneling charge source for ~~the~~ a floating gate of the memory transistor ~~404~~ 400. The drain implant region ~~402B~~ 402 and the buried ~~implants 404B~~ implant 404 are spaced apart, so as to define an active region ~~414B~~ 414 therebetween. Accordingly, the drain implant region ~~402B~~ 402 connects to [[a]] the bitline BL_i. The source implant region ~~406B~~ 406 connects to [[a]] the source line S_i.

A first poly layer ~~410B~~ 410, forming [[a]] the floating gate of the memory ~~cell~~ transistor 304 ~~floating gate~~, ~~overlies~~ overlays the buried implant region ~~404B~~ 404, separated therefrom by a gate ONO layer ~~450B~~ 450. A second poly layer ~~408B~~ 408, forming a common control gate, extends continuously over ~~floating poly~~ the first poly layer 410 (which forms the floating gate) from the source dopant region ~~406B~~ 406 to the drain dopant region 402, ~~overlying~~ overlaying both the ~~floating gate region~~

~~404B~~ buried implant 404 and the select transistor 302 active region ~~414B~~ 414. A tunnel oxide ~~460B~~ 460 of thickness 50-70 angstroms is formed in a tunnel window region between the buried implant ~~404B~~ 404 and the floating gate 410B.

~~The~~ An exemplary manufacturing process of the memory cell ~~400B~~ 310 is shown in the flowchart of Fig. 6 and ~~the~~ a result after each step is shown in Figs. 5A-5G. With reference to Fig. 5A, according to a preferred process of manufacturing the present invention, at step 602, a screen oxide 504 is deposited over ~~the~~ a substrate 502. The thickness of the screen oxide layer is approximately 150 angstroms.

Referring to Figs. 5B and 6, at step 604, a photoresist mask 506 is applied at face 504 of substrate 502. This mask 506 is patterned so as to permit ion implantation of a floating gate region ~~508~~ though gaps in the photoresist mask 506. ~~Next, a~~ A buried N+ tunnel region 508 is implanted in semiconductor substrate 502 through the opening of the mask 506 and the mask 506 is then removed using a conventional process. The substrate 502 is then annealed in, for example, a ~~900° C~~ 900 °C nitrogen environment to ameliorate damage caused to substrate 502 by the prior implantation step 604 and to diffuse the tunnel implant region 508 into substrate 502.

Referring to Figs. 5C and 6, at step 608, after the annealing treatment of the substrate 502, another mask 510 is formed on top of the oxide layer 504 for memory cell implantation. ~~The resulting~~ Resulting cell implant regions 514 and 516 and buried implant region 512 are seen in Fig. 5D.

Referring to Figs. 5D and 6, at step 610, the screen oxide is etched away and an initial gate oxide layer 517 is formed in its place.

Referring to Figs. 5E and 6, at step 612, a tunnel window mask 530 is deposited to a very high

thickness so that ~~the~~ an opening for a tunnel oxide layer ~~510~~ 513 can be precisely positioned at the openings of this tunnel window mask 530 layer above the buried implant 512.

With reference to Figs. 5F and 6, at step 614, after etching away the gate oxide layer 517 in the tunnel windows, ~~the~~ a thin tunnel oxide layer 518 is deposited to a thickness of about 50-70 angstroms. In ~~the~~ a preferred embodiment the tunnel oxide layer 518 represents a thin, high quality silicon dioxide layer which may either be grown in a dry O_2 and HCl mixture atmosphere at a temperature of around 800° to ~~850° C~~ 850 °C. Once the tunnel oxide 518 has been formed, polysilicon floating gates 520 are formed over the gate 517 and tunnel oxide 518 layers ~~517 and 518~~.

Referring to Figs. 5G and 6, at step ~~614~~ 616, an oxide or oxide nitride oxide (ONO) interpoly dielectric 521 is deposited and an etch is performed to create interpoly insulation.

Next, ~~the~~ a control gate poly layer ~~522~~ (not shown) is applied using an LPCVD process. The deposition of the poly layer ~~522 represents~~ is a low temperature application, preferably at less than ~~625° C~~ 625 °C, which tends to maintain [[to]] the poly layer ~~522~~ in an amorphous state.

~~Thus, the process of the present invention next patterns and etches~~ The poly ~~layers 524~~ layer is patterned and etched to produce strips of materials which form control gates 524. The control gate 524 polysilicon ~~522~~ extends beyond the area above the floating gate 520 to adjacent areas to form a common select gate. In addition, this pattern and etch step removes material from the poly layer thereby forming the remaining two sides for each of floating gates 520.

Finally, finishing step ~~616~~ 618 is shown in Fig. 6, such as adding select transistor drain implants

528 and a nitride overcoat may be performed to complete the process. Using the control gate 524 poly layer ~~522~~ as a mask, source implants 528 for the select transistor are made just past the outer edge of the control gate 524 poly ~~522~~.

A memory device constructed according to the teaching of the present invention may be block erased and programmed, and also bit programmed. Referring to Table 1 and Fig. 3, in block programming, memory cell transistor sources, S_0 to S_N, in a block, and also the select transistor drains (the bitlines BL_0 to BL_N) are held at a large negative potential, such as -10 volts, while the memory cell transistor control gates in the block (the wordlines WL_0 to WL_N) are raised to a relatively high positive voltage, such as 10 volts. This causes tunneling of electrons from the buried implant through the tunnel oxide ~~onto~~ into the floating gates ~~512~~.

Memory cells may be block erased by leaving sources S_0 to S_N in the block open, and reversing the word and bitline voltages from the block programming case. Placing bitline electrodes in the block at a relatively high positive voltage, such as 10 volts, and the wordline electrodes in the block at negative 10 volts, causes electrons be expelled out of the floating gate region ~~512~~ back into the buried implant.

Bit programming involves applying a large positive potential to the wordlines and to all bitlines except a selected bitline BL_{i+1}, which is [[a]] at ground potential. The source lines S_0 to S_N are left open.

Memory cells in the present invention may be read by placing the control gate WL_{i+1} of ~~the~~ a particular cell ($_{i+1}$) to be read at positive V_D, and at the same time, placing the drain (bitline) of the particular cell to be read at a relatively low (about 1 volt) voltage V_s. All source lines S_0 to S_N are grounded in read mode. Cells

not in the selected word (row) and bit column have negative[[s]] V_D ~~voltage~~ voltages applied to their wordlines and bitlines that are open.

	WL_i	WL_{i+1}	S_i	BL_i	S_{i+1}	BL_{i+1}	S_{i+2}	BL_{i+2}
Block programming	+10V	+10V	-10V	-10V	-10V	-10V	-10V	-10V
Block Erase	-10V	-10V	Open	+10V	Open	+10V	Open	+10V
Bit (i+1) program	+10V	+10V	Open	+10V	Open	OV	Open	+10V
Read (I+1)	-VD	VD	GND	Open	GND	Vs~1V	GND	Open

Table 1: Voltages Required for Block Programming/Erasing in a Block Alterable Memory.

~~With reference to Table 1 at the end of this specification, in order to achieve block alterable memory, the memory cell 110 in the flash array 100 as shown in Fig. 1 needs to apply +10 volts or -10 across the wordline WL_1 102, the source line S_1 104, and the bitline BL_1 106. Accordingly, the placement of such high voltages to a single memory cell transistor 110 presents reliability and durability problems. Over long periods of time, placing high voltages on the memory device 100 may alter a program stored in each cell 110.~~